**Information about the exam:**

Just wanted to send you some more information about what to expect in the exam and how to prepare, and to answer some questions that I received.

**Past exam questions:**

I have now shared the solutions to the past exam questions on Canvas. I would recommend that you first answer all the questions yourselves and then compare your answers against the solutions. If something is not clear about the solutions, you can ask Tamara and Oliviero at the Q & A session on Thursday, 11am.

**Topics covered:**

As I mentioned on Friday, the guest lectures are NOT included in the exam. All other lectures are examinable.

Key topics / methods:

* Morphology and morphological processing (morphemes, stemming, lemmatization)
* Language models (n-gram language models)
* Part-of-speech tagging (with Hidden Markov Models)
* Syntax, formal grammars and syntactic parsing (simple chart parsing discussed in the lecture, PCFGs and their parameter estimation)
* Distributional semantics (representing meaning through contextual use, count-based models, PMI, clustering)
* Dense word representations (skip-gram model)
* Compositional distributional semantics (lexical function models)
* Neural sentence representations (recurrent neural networks, LSTMs)
* Discourse processing (discourse relations, co-reference resolution, neural document encoders)
* Summarisation

I gave a course overview presentation last Friday. If you were not there, take a look at the slides: <https://cl-illc.github.io/nlp1-2020/resources/slides/NLP1-lecture14.pdf>

**How much to know about a particular task / method:**

If I discussed a method in detail in the lecture, then you need to know it in detail. If I discussed a method only briefly at a conceptual level, then you only need to understand what it does at a conceptual level.

Since this is an introductory NLP course, its aim is to give you a broad overview of the field, i.e. what kinds of problems and tasks there are and what kinds of methods have been proposed for them. So inevitably many of the techniques have only been discussed briefly. Please do make sure, however, that you understand all of the concepts mentioned in the slides. Video lectures and Jurafsky & Martin provide good reference material.

**Types of questions to expect:**

The pen-and-paper exercises should give a good idea of what kinds of questions you may expect in the exam. For instance, there might be questions of the following types:

* Explain a particular linguistic phenomenon and why it is challenging for particular NLP methods / applications
* Explain a particular method and how it is applied; its strengths / limitations (to the extent discussed in the lectures). If you need to give the formulae, this will be stated.
* Apply a method to a given example
* Given examples of system errors, explain why these arise
* Integrating material from different parts of the course: How can one apply a method from one NLP task to solve a particular problem in another NLP task (here, you will need to reflect and write down your suggestions)

**What is allowed at the exam:**

The exam is open-book. You are allowed to use the slides and Jurafsky & Martin, as well as a calculator and an English dictionary. You are NOT allowed to communicate with anyone other than the examiner (i.e. the lecturer or TA present at your zoom session), not in person and also not via electronic communication channels.

**Exam logistics**

We have posted more information about the logistics of the exam on Canvas (see the Exam Logistics page). Please read this information carefully and follow all the steps.